

## PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2001-282456

(43)Date of publication of application : 12.10.2001

(51)Int.Cl. G06F 3/033

G06F 3/00

G06F 3/03

G06K 7/00

G06T 1/00

G06T 7/00

(21)Application number : 2000-101513

(71)Applicant : JAPAN SCIENCE &  
TECHNOLOGY CORP

(22)Date of filing : 03.04.2000

(72)Inventor : KOIKE HIDEKI  
SATO YOICHI  
KOBAYASHI TAKAKUNI

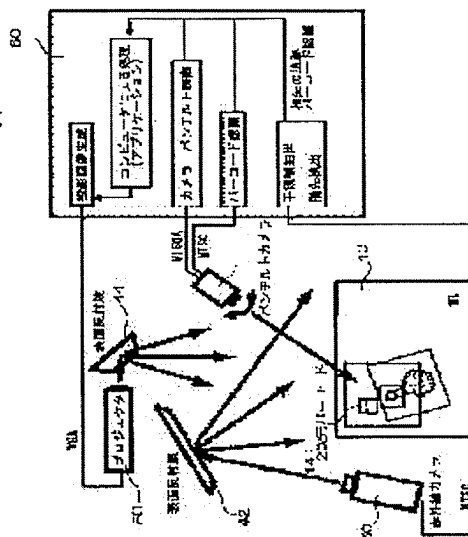
### (54) MAN-MACHINE INTERFACE SYSTEM

#### (57)Abstract:

PROBLEM TO BE SOLVED: To provide a man-machine interface system to use an image of a fingertip on a desk.

SOLUTION: A processing to be performed by calculating a coordinate of the fingertip in a system 60 is explained.

1. An area of a user's hand placed on the desk 10 is extracted from the image of an infrared camera. 2. A fingertip point of a user is detected from the extracted hand area. 3. A position coordinate of the fingertip point on a camera image is transformed into a position coordinate on the desk by calibration by using a preliminarily calculated projective transformation parameter. 4. Interaction is taken with an object and information projected on the desk based on fingertip position information on the desk.



### LEGAL STATUS

[Date of request for examination]	20.08.2003
[Date of sending the examiner's decision of rejection]	
[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]	
[Date of final disposal for application]	
[Patent number]	3834766
[Date of registration]	04.08.2006
[Number of appeal against examiner's decision of rejection]	
[Date of requesting appeal against examiner's decision of rejection]	
[Date of extinction of right]	

*English translation of D3***\* NOTICES \***

JPO and INPIT are not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.\*\*\* shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

**CLAIMS****[Claim(s)]**

[Claim 1] The infrared camera which is a MAMMA scene interface system using the image of a paper fingertip, and obtains the image on a desk, The image by said infrared camera is inputted and it has the processing system to process. Said processing system The MAMMA scene interface system characterized by searching for the coordinate on the desk of the fingertip which specified with pattern matching with the pattern which prepared the fingertip beforehand from the hand field which extracted the hand field using temperature and was extracted, and was specified from the image.

[Claim 2] It is the MAMMA scene interface system characterized by said pattern prepared beforehand being a circular pattern in a MAMMA scene interface system according to claim 1.

[Claim 3] It is the MAMMA scene interface system characterized by including the processing recognized to be pointing actuation when the number of the fingertips which asked for said processing system further in the MAMMA scene interface system according to claim 1 or 2 is one.

[Claim 4] It is the MAMMA scene interface system characterized by performing control which doubles the camera station of said camera with the coordinate on the desk of the fingertip when it has been recognized as said processing system being said pointing actuation, while having further the camera which can perform control of a camera station in a MAMMA scene interface system according to claim 3.

[Claim 5] It is the MAMMA scene interface system which said processing system captures the image photoed with said camera in said MAMMA scene interface system according to claim 4, and is characterized by extracting a bar code and performing bar code recognition processing from the captured image.

[Claim 6] It is the MAMMA scene interface system characterized by for said processing system inputting voice from said voice input means while having a voice input means further in said MAMMA scene interface system according to claim 1 to 5, and performing speech recognition processing.

[Claim 7] The MAMMA scene interface system which is equipped with a projector and is further characterized by projecting a computer screen on a desk by said projector in a MAMMA scene interface system according to claim 1 to 6.

[Claim 8] The record medium which stored the program to which the processing performed with a MAMMA scene interface system according to claim 1 to 7 is made to carry out to computer system.

[Translation done.]

**\* NOTICES \***

JPO and INPIT are not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.\*\*\*\* shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

---

**DETAILED DESCRIPTION**

---

[Detailed Description of the Invention]

[0001]

[Field of the Invention] About a MAMMA scene interface, especially, this invention does not make use of a computer conscious of the existing actuation, but thinks a user's amenity as important, and relates to the interface to which a user's degree of freedom is not reduced by attaching a measuring instrument etc. in a user's body.

[0002]

[Background of the Invention] A computer permeates our life deeply and the use gestalt is also various. Before, a very small computer which was not considered is carried or use in a location which is embedded in a surrounding environment, glances and does not turn out to be it is also performed. It is thought that it meets and develops into two flow, the gestalt attached in a man like a wear rubble personal computer or PDA in the future and the gestalt united with the environment around [, such as a desk, and a wall, a room, ] a user. When the MAMMA scene interface was considered in such flow, the interface represented by the present GUI etc. had to bind the user to the interface, and the interface also had to be taken care of apart from the activity to carry out originally. So, in the next-generation interface, the interface which can perform more direct and more natural actuation to the activity which a user wants to do is searched for.

[0003] Then, in the informational retrieval and the use using a computer, for mitigation of the burden by a user's interface, a computer supports an activity and actuation of a user in the real world, and it has a hope for real world orientation and the interface by situation recognition as an interface of the next generation which can be used more comfortably than the interface of the conventional type only by the keyboard or the mouse. As for a real world orientation user interface, a computer always recognizes the situation of the user in the real world, an intention of a user is drawn up, and a computer supports an activity. This aims at implementation of the transparent interface in which actuation of the existing computer is not impressed. This real world orientation interface is applied to a paper activity, and development of the real world orientation interface (Enhanced Desk) integrated by the "desk" is furthered as an example of a \*\*\*\*\* world orientation interface. As an example which took notice of fusion of a paper activity and an activity by computer, Digital Desk (for example, pool "Bit separate volume visual aiming at interface-post GUI -" refer to KYORITSU SHUPPAN, Chapter 2.1, and pp.24-44 grade) is known well. At Digital Desk, the computer screen projected on the desk is operated by a fingertip etc. A user can cut off and copy the illustration projected on the desk, or can calculate with the calculator projected on the desk.

[0004] Paying attention to the paper document on a desk, there is also an example which has tried integrated use of a paper document and electronic intelligence (M. Kobayashi and H.Koike: Enhanced Desk, Integrating Paper Documents and Digital Documents; Proceedings of 1998 Asia Pacific Computer Human Interaction, and pp.167-174 reference (1998)). Although the bar code

beforehand given to the paper document for matching of a paper document and electronic intelligence was used in this example, since it was required to observe a bar code in the above magnitude to some extent, the large-sized bar code had to be used. Moreover, in order to recognize a user's hand field, the beige extract was performed, but when it will have been incorrect-recognized if the near body was beige on the desk, or the image was projected on a user's hand, there was a trouble that a hand field extract did not work. Always assuming the sense of a specific hand also about recognition of a user's fingertip location, the fingertip was the problem with much [ and ] constraint of considering as one point that the insufficiency of recognition precision was big. And since software was performing a series of processings, real-time processing was impossible.

[0005]

[Problem(s) to be Solved by the Invention] The purpose of this invention is information interface system implementation integrated by the "desk" as an interface united with the environment around a user. In office, the activity using PC is done on a desk, and the activity using a paper document is done on coincidence in many cases. The information interface system distribution which can treat the activity on these desks integrative by making a computer observe a desk top is performed.

[0006]

[Means for Solving the Problem] The infrared camera which this invention is a MAMMA scene interface system which uses the image of a paper fingertip, and obtains the image on a desk in order to attain the above-mentioned purpose. The image by said infrared camera is inputted and it has the processing system to process. Said processing system It is the MAMMA scene interface system characterized by searching for the coordinate on the desk of the fingertip which specified with pattern matching with the pattern which prepared the fingertip beforehand, and was specified from the hand field which extracted and extracted the hand field from the image using temperature. Said pattern prepared beforehand can also be used as a circular pattern. Thereby, the recognition precision of a fingertip can be raised and, moreover, recognition processing can be performed at a high speed.

[0007] Furthermore, when the number of said fingertips for which it asked is one, it can be recognized as it being pointing actuation, and it is this pointing actuation and pointing actuation performed with a mouse etc. can be performed. Furthermore, while having the camera which can perform control of a camera station, control which doubles the camera station of said camera with the coordinate on the desk of the fingertip can be performed by performing said pointing actuation. Thereby, the image photoed with said camera is captured, from the captured image, a bar code can be extracted and bar code recognition processing can also be performed. Furthermore, while having a voice input means, a command with voice etc. can be inputted by inputting the voice by said voice input means, and performing speech recognition processing. Furthermore, said projector can also perform interactive processing with a computer screen by projecting a computer screen on a desk by having a projector. The record medium which stored the program to which above-mentioned processing is made to carry out is also this invention.

[0008]

[Embodiment of the Invention] The MAMMA scene interface system of this invention has the following descriptions.

- (1) Extract a user's hand field to stability by measuring the synchrotron orbital radiation from a user's skin field using an infrared camera.
- (2) The use of template matching based on normalization correlation etc. realizes a high-speed trace of a user's fingertip.
- (3) In order to recognize a paper small bar code, carry out the expansion trace of a user's fingertip circumference using a pan tilt camera with a zoom function.
- (4) Perform use of image-processing hardware, and application of distributed processing for improvement in the speed of processing.

Below, the operation gestalt of this invention is explained with reference to a drawing focusing on these descriptions at a detail.

[0009] Drawing 1 is the external view of the enhanced strike disk (Enhanced Desk) which is the operation gestalt of this invention. As shown in drawing 1, a projector 30 is attached above the usual desk 10, the screen of a computer (not shown) is reflected in a mirror 44, and Screen 12 is projected on a desk. Two cameras which observe a desk top were installed. One set is an infrared camera 50, turns and installs a top from desk 10 left-hand side, and performs paper photography throughout a period of surface reflecting mirror 42. One more pan tilt camera 20 with a zoom function can be installed in the upper right direction of a desk, and can expand and photo a paper part. Moreover, although not illustrated, the loudspeaker and the microphone are also installed and it also has the function of I/O, such as voice.

[0010] Drawing 2 shows the system configuration Fig. in an operation gestalt. In drawing 2, the projection image generated with computer system 60 is projected by the projector 30 on a desk 10 through the surface reflecting mirror 44. Moreover, the posture of the pan tilt camera 20 which is picturizing the image on a desk 10 etc. is controlled by computer system 60. The image picturized with the pan tilt camera 20 is also captured and processed in computer system 60. The image on the desk picturized with the infrared camera 50 is also captured and processed by computer system 60. The flow of the main processings with the enhanced strike disk of this invention is shown below.

\*\* Extract a user's hand field from the image of an infrared camera 50.

\*\* Detect a user's fingertip point from the extracted hand field.

\*\* Change into a paper position coordinate from the position coordinate of the fingertip point on a camera image using the projective-transformation parameter for which it asked beforehand by the calibration.

\*\* Take the object and information which were projected on the desk based on paper fingertip positional information, and interaction.

These processings are explained to a detail below.

[0011] the extract of a <extract of hand field> hand field — a background — generally the beige extract with difference or a color camera is used. However, it is difficult to extract a hand field by such technique depending on the condition of a background in many cases. In the work environment assumed especially by this system, in order it is possible to open a paper document, books, etc. and to project electronic intelligence by the projector on a desk, there is a trouble that the color of a user's hand field is not fixed etc. So, in this system, in order to extract a hand field to stability, the infrared camera 50 was used. From the image of the infrared camera which is observing the desk top, a hand field can be started based on people's temperature. The situation of a paper hand can be obtained as an image of 256 gradation of NTSC using an infrared camera by taking a photograph near people's temperature (30 degrees C – 34 degrees C). This image is binary-ization-processed with a fixed threshold, and a hand field is extracted. Extract processing of a hand field is explained to drawing 3 based on the image from an infrared camera. Drawing 3 (a) is an image from an infrared camera 50. The resolution at the time of incorporating with computer system 60 is 256x220 pixels. And the image near people's temperature (30 degrees C – 34 degrees C) is extracted from this image. Drawing 3 (b) showed this, and it extracts and makes the part of a hand field binary. As shown in drawing 3 (b), it is not influenced by change of a background or lighting by this technique, but only a hand field can be extracted to stability. For example, NIKON thermal BISHON (LAIRD3A) can be used for an infrared camera. This camera can photo the range to -20 degrees C – 50 degrees C in field time 1 / [ 768x465 pixels of effective pixel numbers, and ] 60 seconds.

[0012] detection of a <recognition of fingertip location> fingertip location has a user's hand on a desk, and the distance of the hand of a camera and a user can detect a fingertip location by the template of fixed magnitude from seeing by about 1 law and the upper magnitude not changing extremely. Generally recognition of a fingertip location is detected using a circular template

based on the profile of a fingertip configuration being close to a circle by performing pattern matching based on normalization correlation around a hand field. A user's hand is on a desk, from the magnitude on appearance not changing extremely, the object of pattern-matching processing is limited only within the limits of [ fixed ] a hand field, and improvement in the speed of processing is attained. Recognition of this fingertip location is explained using drawing 4 . Pattern-matching processing performs coincidence detection using a correlation value using a template with a magnitude of 15x15 pixels shown in drawing 4 (a) to the 256x220-pixel image (refer to drawing 3 (b)) which started and made for example, the hand field binary. When many points that a correlation value is high are found after pattern-matching processing in the location which approached too much, it leaves the point of them that a correlation value is the highest, and the remainder is excepted from a fingertip candidate point. Since some fingertip candidate points are acquired by this, the point of having been incorrect-recognized other than a fingertip point is excepted. As shown in drawing 4 (b), this processing investigates the pixel corresponding to eight perimeters (four rectangular top-most vertices and point of each side dividing into two) of a template, and judges whether it is a fingertip point. When the fingertip has run through the template by this processing or anything does not have a pixel appropriate for a finger in (ERROR of drawing 4 (b)), and the perimeter of a template, it excepts from a fingertip point candidate. Finally, as shown in drawing 4 (c), a fingertip candidate point is adopted as order with a high correlation value to five, and it considers as a user's fingertip point. It is carrying out using the Hitachi image-processing board IP 5010 as an example of the fingertip detection processing in this system. IP5010 is the image-processing board which had the memory for 40 screens by monochrome shade image, and is the board which can perform binary-izing and an image processing called difference at a high speed by hardware between the images saved on memory. OS supports WindowsNT (trademark) and Linux and many image-processing functions are prepared as a library. In this example, Linux was used for OS in view of compatibility with a network.

[0013] In order to know whether the location detected in the <amendment of camera image coordinate and desk Kamitaira side coordinate> infrared image is equivalent to which location on a desk, amendment of an infrared camera image coordinate and a desk Kamitaira side coordinate is performed. Here, some corresponding points of an infrared camera image coordinate and a desk Kamitaira side coordinate are defined, and correspondence between an infrared camera image coordinate (x y) and a desk Kamitaira side coordinate (x', y') is used and expressed for projective transformation like a formula (1) based on the relation.

[Equation 1]

$$\begin{cases} x' = c_1x + c_2y + c_3xy + c_4 \\ y' = c_5x + c_6y + c_7xy + c_8 \end{cases} \quad (1)$$

In order to ask for corresponding points, it explains using drawing 5 explaining measurement of infrared camera corresponding points. The system of coordinates of a desk Kamitaira side are set up to a desk 10. For this reason, the calibration plate 70 is created in the location of a paper target point, and a desk upper seat label system is set up by placing this on a desk. In the infrared camera which starts a hand field, since a temperature gradient is outputted as an image, as shown in drawing 5 , a miniature lamp 72 is embedded at the corresponding points of this calibration plate, the turned-on electric bulb is photoed with an infrared camera, and corresponding points are measured. A projective-transformation parameter (c1-c8) is determined based on the relation after measuring the group of corresponding points. The distortion of an image, a gap of a location, etc. can be amended by changing using this parameter. In order to obtain this projective-transformation parameter, there should just be at least 4 sets of corresponding points, but in this system, in order to ask stability for a solution,

as shown in drawing 5 , 9 sets of corresponding points (miniature lamp) were prepared. It asked for the parameter of conversion from these 9 sets by solving the above-mentioned simultaneous equations using a singular-value-decomposition method. In this system, paper system of coordinates and the system of coordinates of the image projected on the desk by the projector are made the same for processing mitigation, and the position coordinate of the value of a desk Kamitaira side coordinate and the pixel of a projector projection image is made equal. Thereby, in order to project the image of a projector on the fingertip location on a desk Kamitaira side, mitigation of processing can be aimed at that what is necessary is just to only project the image which drew the object on the coordinate value on a desk Kamitaira side coordinate.

[0014] In a <bar code recognition by gaze of the fingertip circumference> book system, when the number of the fingertips of the user who has recognized by performing an image processing is one, a user can also judge with carrying out pointing actuation. It can use instead of pointing, such as a mouse, by recognizing it as it being pointing actuation by pointing out only by the index finger, and acquiring the coordinate of this fingertip by this. It can also be considered that it is the field which the circumference of the fingertip which is carrying out this pointing actuation is observing. The pan tilt camera 20 which is this application explains below the processing which pursues the attention field of one fingertip.

[0015] (Flow of fingertip trace processing) The flow of processing of the fingertip trace in the pan tilt camera 20 is shown below.

\*\* Measure corresponding points among 2 flat surfaces of a desk Kamitaira side and a pan tilt drive flat surface, and compute the projective-transformation parameter.

\*\* Measure the location of the fingertip point on an infrared camera image by the image processing.

\*\* Change the location of the fingertip point on an infrared camera image into the coordinate on a desk Kamitaira side by projective transformation.

\*\* Use projective transformation and change the location of a paper fingertip point into a pan tilt drive flat surface.

\*\* Send out the coordinate on a pantilt drive flat surface to a camera in VISCA code, and make it turn [ camera ] to a user's fingertip point.

\*\* Process the image of the camera suitable for a fingertip point.

[0016] (Calculation of a projective-transformation parameter) In this system, it should just be turned out whether a paper fingertip location is located on the pan tilt drive flat surface of a camera 20 in which coordinate location. Since it can know which coordinate value the pan tilt camera 20 has taken on the drive flat surface now, this processing is performed using this. Although it explained having performed amendment of a camera image coordinate and a desk upper seat label previously, also in the fingertip trace, coordinate amendment with a camera 20 is performed using the same technique. Therefore, first, since the desk upper seat label of a fingertip is known by old processing, the location on the drive flat surface of the pan tilt camera 20 corresponding to the desk top location is computed. Four or more sets of corresponding points are measured. This measurement adjusts a location so that a pan tilt camera may catch paper corresponding points at the core, and it measures the coordinate value on the pan tilt drive flat surface in that time as corresponding points. Previous projective transformation is used based on these corresponding points, and it asks for the parameter changed into the coordinate on a pan tilt drive flat surface from a desk upper seat label. The camera 20 for a fingertip trace has the function in which a camera can be turned in the direction by specifying a coordinate. A user's fingertip is made to pursue using this function. As an example, the tracking camera 20 for a fingertip trace is perpendicularly. – They are 7794–7794, and a horizontal direction. – Pan tilt control of a camera is performed by using a Sony EVI-G20 pan tilt camera with the pan tilt drive flat surface of the magnitude of 15570–15570, and transmitting the VISCA command to a camera from the serial port of PC.



[0017] (Fingertip trace processing) Now, fingertip trace processing is explained. By the above-mentioned image processing, the location of the fingertip point on a desk 10 is measured for the image from an infrared camera 50. While changing the location of the fingertip point on an infrared camera image into the coordinate on a desk Kamitaira side by projective transformation, it recognizes that it is only one fingertip. With an above-mentioned parameter, projective transformation is used and the location of the fingertip point on the desk for which it asked is changed into the pan tilt drive flat surface of a camera 20. The coordinate on a pan tilt drive flat surface is sent out to a camera in VISCA code, and it is made to turn [ camera / 20 ] to a user's fingertip point. Thereby, the image of the camera 20 which turned to the fingertip point can be captured.

[0018] (Recognition of a bar code) The image of this camera 20 is used for below, an expansion trace of a fingertip is performed to it, and the example which recognizes small bodies, such as a bar code, is explained to it using drawing 6. Drawing 6 performs an expansion trace of a fingertip and shows signs that bar code recognition is performed. On actual bodies (a paper document, books, etc.), it can recognize by attaching a bar code, and a link with electronic intelligence can be created on them. The 2-dimensional matrix code was used for the bar code. This two dimensions bar code can calculate an objective posture from a code, if the location not only with the class of code but a code and the location where the objective code is stuck since the sense etc. can be obtained from a code are memorized. The example of a two-dimensional bar code is shown in drawing 6 (a). Such a bar code will be recognized if it exists in the image photoed with the pan tilt camera 20. In order to recognize, the magnitude of extent which is among a camera image is required, but since the image which the pan tilt camera 20 expanded is photoed, it is possible to recognize to a thing with a magnitude of about 1.5x1.5cm also at the bar code stuck on the document set on the desk. On a desk, by pointing actuation of a user, if a bar code is pointed at, the code can be recognized and the interaction corresponding to it can be raised. Signs to this bar code that the expansion trace of a fingertip was performed are shown in drawing 6 (b). Drawing 6 (b) shows the image which made the image of a camera 20 binary. First, the location of the fingertip point on the image from an infrared camera 50 is measured by the image processing. Thereby, the position coordinate of having pointed out with one finger and its fingertip point can be acquired. Next, the location of the fingertip point on an infrared camera image is changed into the coordinate on a desk Kamitaira side by projective transformation. Projective transformation is used for a pan tilt drive flat surface, the location of the fingertip point on the obtained desk is changed, the coordinate on the acquired pan tilt drive flat surface is sent out to a camera 20 in VISCA code, and it is made to turn [ camera / 20 ] to a user's fingertip point. And the image ( drawing 6 (b)) of the camera 20 which turned to the fingertip point is processed, a bar code image is recognized, and reading processing of a bar code is performed. Recognition processing of a bar code can incorporate for example, a video image in the magnitude of 320x240 pixels as an example from the video input terminal of O2 which is the system of SGI, can process binary-ization etc., and can be recognized using a bar code recognition library. The image input is using the video library of SGI. Moreover, software is performing bar code recognition processing and it can perform at the rate of 10 - 15 f/s extent.

[0019] Speech recognition is used for a <speech recognition> book system as an auxiliary role of interaction with a user. Here, it uses only in order to make the keyword for changing the operation modes, such as pointing actuation of a user, migration of an object, and rotation, recognize. Speech recognition is performed using the speech recognition engine of IBM ViaVoice. In this system, when a user utters the word registered beforehand, in order to raise a certain interaction, the speech recognition process has always sent the class of recognized word as a message to the message server. As for the message sent to the server, it is possible for an information presentation process to raise the interaction corresponding to reception and its word on a desk.

[0020] In the <distributed-processing> book system, it roughly divides and four processings of fingertip detection and the fingertip trace processing with a pan tilt camera, two-dimensional bar code recognition processing, speech recognition processing, and information presentation processing are performed. In this system, the machine of the plurality for improvement in the speed is performing distributed processing for each processing. The schematic diagram of distributed processing is shown in drawing 7. Fingertip detection and the fingertip trace processing with the pan tilt camera 20 are for example, Pentium(trademark) II. It is carrying out using the machine which introduced Linux and the image-processing board IP 5010 into the personal computer 64 of 450Mhz. According to this process, a user's paper number of fingertips and paper location were measured, and the desk upper seat label location of a fingertip is sent out to the message server 68. Moreover, the camera 20 is controlled to become a fingertip location on a user's desk about the direction of a look of the pan tilt camera 20. By two-dimensional bar code recognition processing, it is SGI about the image of the pan tilt camera 20. It incorporates to the video input of O2 system 66, and the two-dimensional bar code is recognized by software processing. A recognition result is always sent out to the message server 68. Speech recognition is processing by introducing WindowsNT and ViaVoice into the personal computer 62 of PentiumII 330Mhz. When the word which is made to recognize all the voice about which the user spoke toward the microphone, among those has been registered beforehand has been recognized, the specific message is sent out to the message server 68. Finally the information always stored in the message server 68 is taken out, and there is an information-display process which creates the interaction and image corresponding to it. So to speak, this is an application process, and is SGI. It is processing by O2 system 66. In the message server 68 which performs the communication link during each processing, it is tuple space communication system TS. System/s and a common name Linda are used. This system 68 has realized the communication link by sharing the space called a tuple tooth space on a network, and exchanging the message called the tuple which considered the character string of arbitration as the set through that space. Since it is asynchronous communication, each process can operate independently completely and there is an advantage that an addition and modification of a new process are easy. In this system, at the practical rate of per second 20 or more frames, it was stabilized and the fingertip was able to be detected.

[0021] A user can recognize the alphabetic character written with the finger using pursuing pointing actuation of the user on a <example of application> (simple character recognition) desk. For example, if a figure is written in from a projector 30 with a finger to the grid projected on the desk, the figure can be recognized and the figure which wrote can be displayed by the projector 30 on a desk. It can use in order to give semantics other than simple pointing to a user's fingertip.

(Other application) As other application, not only the positional information of a user's fingertip but the gesture recognition using the locus of five fingertips etc. is possible. Since the photomacrography with a camera became possible and it is thought that the small alphabetic character of space can also be recognized, character recognition is used, and although the function for integrative use of a paper document and electronic intelligence is added, it can do. Moreover, construction of the communication environment through the desk Kamitaira side by sharing the two desk top of a remote place for this system as application to a remote coordination activity can also be performed. This invention may be applied to the client/server system which consists of not only an above-mentioned distributed computer system but two or more systems, or a stand-alone system. The configuration of this invention is realizable by reading and performing a program by the system from the storage which stored the program about this invention. There are a floppy (trademark) disk, CD-ROM, a magnetic tape, a ROM cassette, etc. in this record medium.

[0022]

[Effect of the Invention] As mentioned above, if this invention is used, an interface required for

the desk "Enhanced Desk" implementation which can treat the electronic intelligence and the actual body on a desk integrative is realizable. Since the expansion trace of the fingertip was carried out with the camera, recognition of a small bar code was attained and it became unnecessary moreover, to attach a big bar code to the body on a desk unnaturally. As for a user, it is more nearly intuitive than a conventional mouse and a conventional keyboard to move, rotating 3D object projected for example, on the desk by using the interface of this invention by the fingertip, or to display the homepage related by pointing at the bar code attached to books etc., and he can perform reality and the interaction linked closely between electronic intelligence.

[Translation done.]

**\* NOTICES \***

**JPO and INPIT are not responsible for any damages caused by the use of this translation.**

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.\*\*\* shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

---

**DESCRIPTION OF DRAWINGS**

---

[Brief Description of the Drawings]

[Drawing 1] It is the external view of an en hunger strike disk.

[Drawing 2] It is the system configuration Fig. of this invention.

[Drawing 3] It is drawing showing the extract of a hand field.

[Drawing 4] It is drawing explaining fingertip recognition.

[Drawing 5] It is drawing explaining measurement of corresponding points.

[Drawing 6] It is drawing explaining an expansion trace of a fingertip.

[Drawing 7] It is drawing explaining distributed processing.

[Description of Notations]

10 Desk

12 Screen

20 Pan Tilt Camera with Zoom Function

30 Projector

42 Surface Reflecting Mirror

44 Surface Reflecting Mirror

50 Infrared Camera

60 Computer System

62 Personal Computer

64 Personal Computer

66 SGI O2 and System

68 Message Server

70 Calibration Plate

72 Miniature Lamp

---

[Translation done.]

\* NOTICES \*

JPO and INPIT are not responsible for any damages caused by the use of this translation.

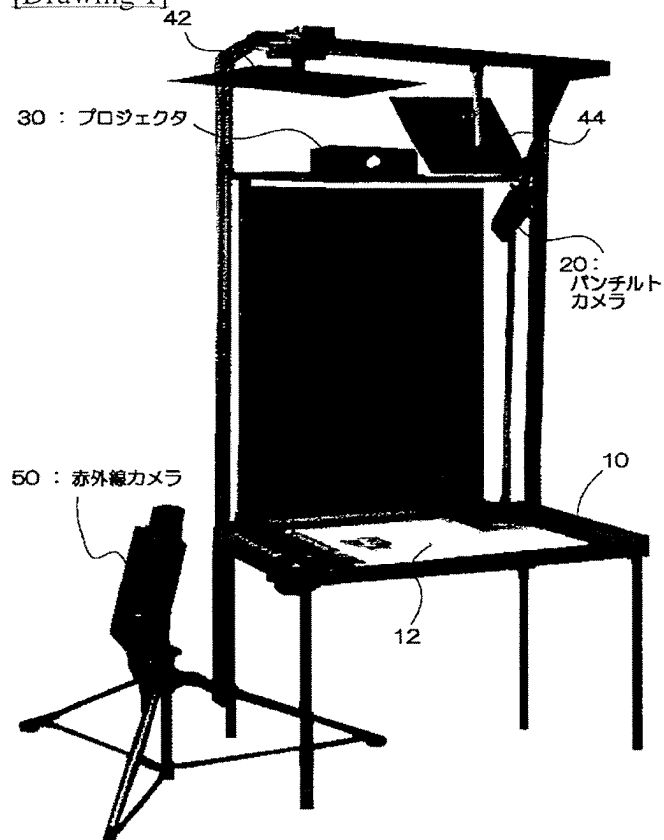
- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.\*\*\*\*\* shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

---

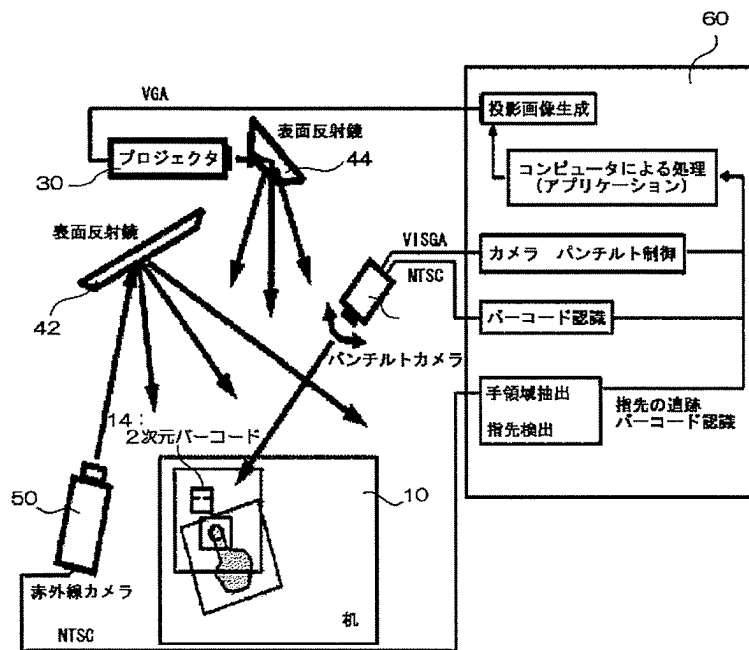
DRAWINGS

---

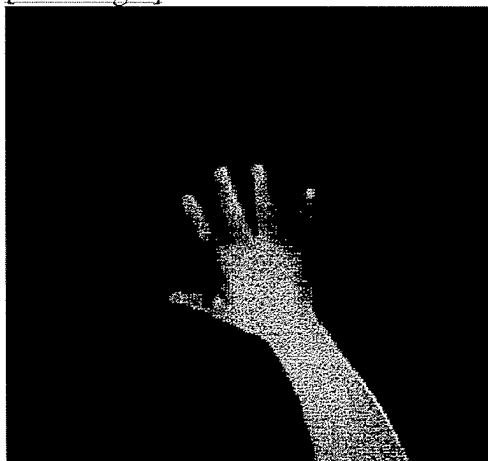
[Drawing 1]



[Drawing 2]



[Drawing 3]

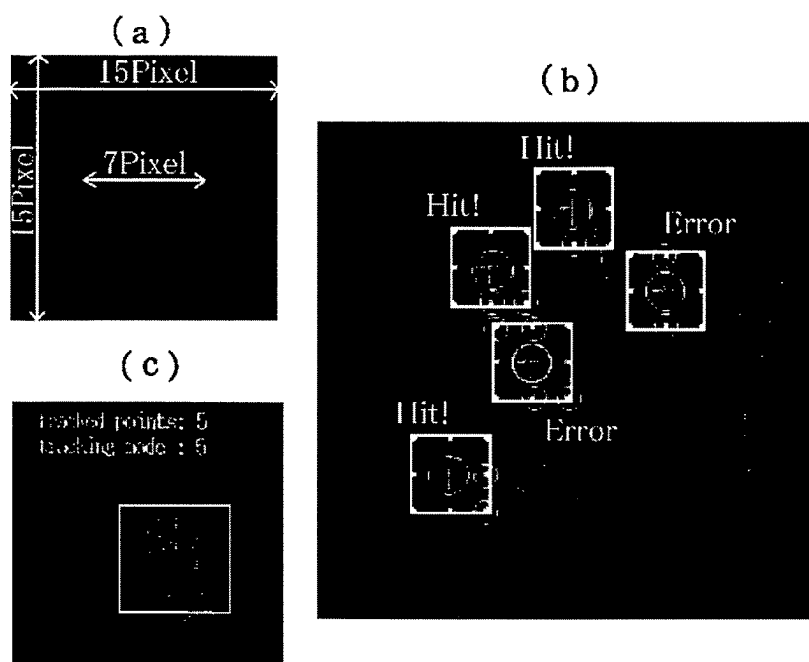


(a) 赤外線画像

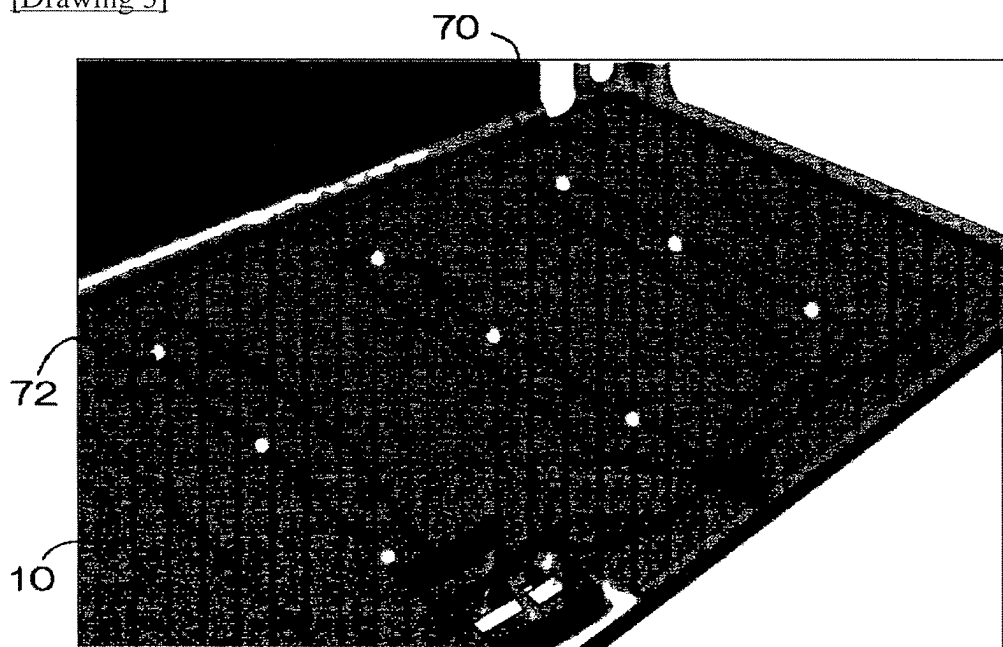


(b) 手領域の抽出

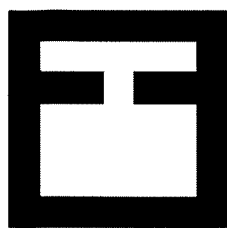
[Drawing 4]



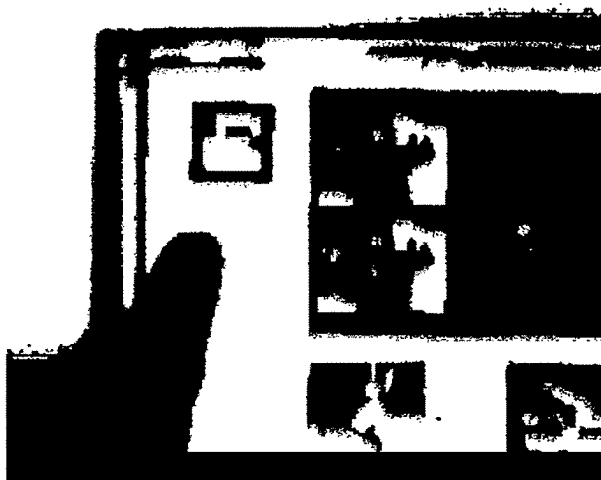
[Drawing 5]



[Drawing 6]

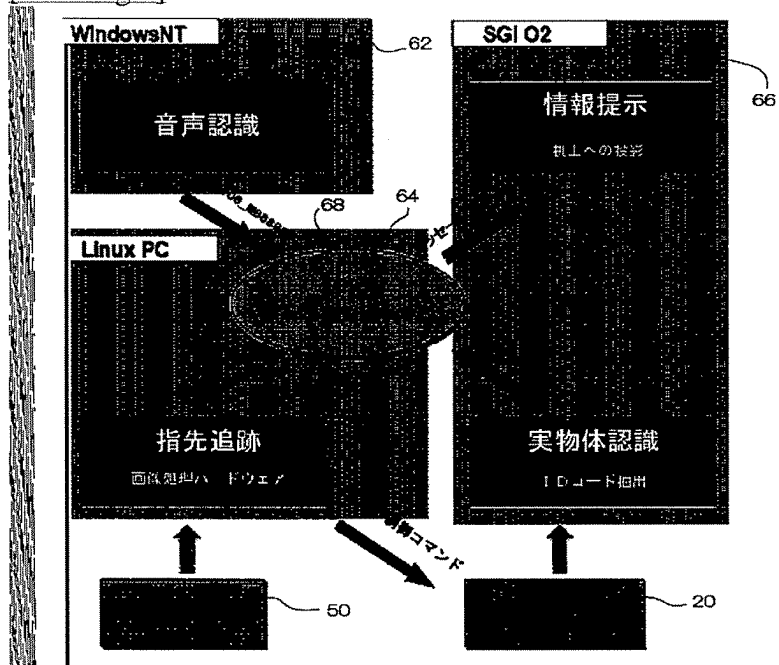


(a)



(b)

[Drawing 7]



[Translation done.]



(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2001-282456

(P2001-282456A)

(43) 公開日 平成13年10月12日 (2001. 10. 12)

(51) Int. Cl. <sup>7</sup>	識別記号	F I	テ-マ-ト*(参考)
G 0 6 F 3/033	3 5 0	G 0 6 F 3/033	3 5 0 G 5 B 0 5 7
3/00	6 3 0	3/00	6 3 0 5 B 0 6 8
	6 8 0		6 8 0 C 5 B 0 7 2
3/03	3 3 0	3/03	3 3 0 Z 5 B 0 8 7
G 0 6 K 7/00		G 0 6 K 7/00	U 5 E 5 0 1

審査請求 未請求 請求項の数 8 O L (全 11 頁) 最終頁に続く

(21) 出願番号 特願2000-101513(P2000-101513)

(22) 出願日 平成12年4月3日 (2000. 4. 3)

特許法第30条第1項適用申請有り (1999年10月4日 ヒューマンインタフェース学会発行の「ヒューマンインタフェースシンポジウム'99論文集」に発表)

(71) 出願人 396020800

科学技術振興事業団

埼玉県川口市本町4丁目1番8号

(72) 発明者 小池 英樹

東京都杉並区荻窪5-29-17-1104

(72) 発明者 佐藤 洋一

東京都江東区東陽5-3-1-501

(72) 発明者 小林 貴訓

神奈川県茅ヶ崎市下寺尾401-1

(74) 代理人 100105371

弁理士 加古 進

最終頁に続く

(54) 【発明の名称】 マンマシーン・インターフェース・システム

(57) 【要約】

【課題】 机上の指先の画像を用いるマンマシーン・インターフェース・システムの提供。

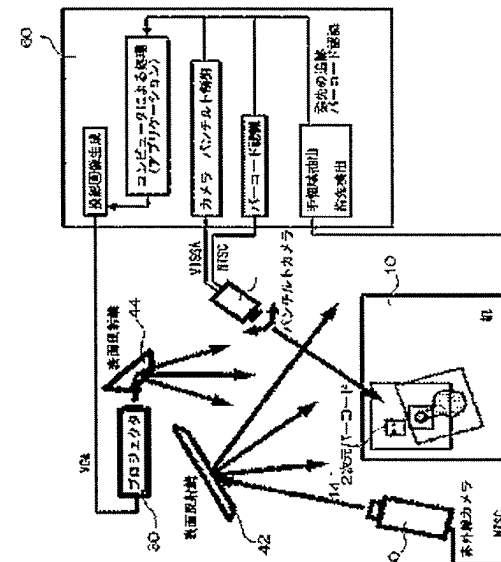
【解決手段】 システム60における、指先の座標を求めて行う処理を説明する。

① 机10上に置かれたユーザの手の領域を、赤外線カメラ40の画像より抽出する。

② 抽出された手領域からユーザの指先点を検出する。

③ カメラ画像上の指先点の位置座標から、キャリブレーションによりあらかじめ求めておいた射影変換パラメータを用いて、机上の位置座標に変換を行なう。

④ 机上の指先位置情報をもとに、机上に投影したオブジェクトや情報とインタラクションをとる。



(2)

特開2001-282456

1

2

## 【特許請求の範囲】

【請求項1】 机上の指先の画像を用いるマンマシーン・インターフェース・システムであって、机上の画像を得る赤外線カメラと、前記赤外線カメラによる画像を入力して、処理する処理システムとを有しており、

前記処理システムは、画像から手領域を体温を利用して抽出し、

抽出した手領域から、指先を予め用意したパターンとのパターン・マッチングにより特定し、

特定した指先の机上の座標を求めることを特徴とするマンマシーン・インターフェース・システム。

【請求項2】 請求項1記載のマンマシーン・インターフェース・システムにおいて、前記予め用意したパターンは、円形のパターンであることを特徴とするマンマシーン・インターフェース・システム。

【請求項3】 請求項1又は2に記載のマンマシーン・インターフェース・システムにおいて、前記処理システムは、さらに、求めた指先が1本であるときは、指差し動作であると認識する処理を含むことを特徴とするマンマシーン・インターフェース・システム。

【請求項4】 請求項3記載のマンマシーン・インターフェース・システムにおいて、さらに、撮影位置の制御ができるカメラを備えるとともに、前記処理システムは、前記指差し動作であると認識されたとき、その指先の机上の座標に、前記カメラの撮影位置を合わせる制御を行うことを特徴とするマンマシーン・インターフェース・システム。

【請求項5】 前記請求項4記載のマンマシーン・インターフェース・システムにおいて、前記処理システムは、前記カメラで撮影した画像を取り込み、

取り込んだ画像から、バーコードを抽出して、バーコード認識処理を行うことを特徴とするマンマシーン・インターフェース・システム。

【請求項6】 前記請求項1～5のいずれかに記載のマンマシーン・インターフェース・システムにおいて、さらに、音声入力手段を備えるとともに、

前記処理システムは、前記音声入力手段から音声を入力し、音声認識処理を行うことを特徴とするマンマシーン・インターフェース・システム。

【請求項7】 請求項1～6のいずれかに記載のマンマシーン・インターフェース・システムにおいて、さらに、プロジェクタを備え、前記プロジェクタにより、コンピュータ画面を机上に投影することを特徴とするマンマシーン・インターフェース・システム。

【請求項8】 請求項1～7のいずれかに記載のマンマシーン・インターフェース・システムで行われる処理をコンピュータ・システムに行わせるプログラムを格納した記録媒体。

## 【発明の詳細な説明】

## 【0001】

【発明の属する技術分野】本発明は、マンマシーン・インターフェースに関し、特に、コンピュータの利用に既存の操作を意識せず、ユーザの快適性を重視し、計測器などをユーザの身体に取り付けることでユーザの自由度を低下させることのないインターフェースに関する。

## 【0002】

【技術的背景】コンピュータは我々の生活に深く浸透し、その利用形態もさまざまである。以前では考えられなかったような非常に小型のコンピュータを携帯したり、周囲の環境に埋め込まれ一見してそれと解らないような場所での利用も行なわれている。将来的にはウェアラブル・パソコンやPDAのような人に取り付ける形態と、机や壁、部屋などユーザの周囲の環境に融合した形態の2つの流れにそって発展してゆくと考えられる。このような流れの中で、マンマシーン・インターフェースについて考えると、現在のGUI等に代表されるインターフェースは、ユーザをインターフェースに縛り付け、本来行ないたい作業とは別にインターフェースにも気を配らなければならなかった。そこで次世代のインターフェースではユーザの行ないたい作業に対してより直接的でより自然な操作を行なうことのできるインターフェースが求められている。

【0003】そこで、コンピュータを用いた情報の検索や利用において、ユーザのインターフェースによる負担の軽減のため、実世界でのユーザの作業や操作をコンピュータが支援し、キーボードやマウスのみによる従来型のインターフェースよりも快適に利用できる次世代のインターフェースとして、実世界指向、状況認識によるインターフェースに期待が寄せられている。実世界指向ユーザ・インターフェースは、実世界のユーザの状況を常にコンピュータが認識し、ユーザの意図を汲み取ってコンピュータが作業を支援する。これにより、既存のコンピュータの操作を感じさせない透明なインターフェースの実現を目指している。この実世界指向インターフェースを机上で作業に応用し、机型実世界指向インターフェースの一例として、“机”に統合された実世界指向インターフェース(Enhanced Desk)の開発が進められている。机上で作業とコンピュータでの作業の融合に注目した例として、Digital Desk(例えば、小池「Brilliant: ビジュアルインターフェース—ポストGUIを目指して—」共立出版、2、1章、pp.24-44等参照)がよく知られている。Digital Deskでは机上に投影されたコンピュータ画面を指先などで操作するものである。ユーザは机上に投影されたイラストを切り取ってコピーしたり、机上に投影された電卓で計算をすることができ、

【0004】机上で紙書類に音目し、紙書類と電子情報の統合利用を試みている例もある(M.Kobayashi and H.Koike: Enhanced Desk. Integrating Paper Documents

(3)

特開2001-282456

3

4

ts and Digital Documents; Proceedings of 1998 Asia Pacific Computer Human Interaction, pp.167-174 (1998)参照)。この例では、紙書類と電子情報の対応づけのためにあらかじめ紙書類に付与したバーコードを用いているが、バーコードがある程度以上の大きさで観察されることが必要なため、大型のバーコードを用いなければならなかった。またユーザの手領域を認識するために顔色抽出を行っていたが、机の上に顔色に近い物体があると誤認識されたり、ユーザの手の上に映像が投影されると、手領域抽出がうまくいかないといった問題点があった。ユーザの指先位置の認識についても常に特定の手の向きを仮定し、指先は1点とするなどの制約が多く、また認識精度の不十分さが大きな問題であった。そして一連の処理をソフトウェアで行なっているため実時間処理が不可能であった。

【0005】

【発明が解決しようとする課題】本発明の目的は、ユーザの周囲の環境に融合したインターフェースとして、“机”に統合された情報インターフェース・システムの実現である。オフィスなどでは、机上でPCを使った作業を行ない、また同時に紙書類を使った作業を行なうことが多い。これらの机上での作業をコンピュータに机上を観察させることによって、統合的に扱うことのできる情報インターフェース・システムの提供を行なう。

【0006】

【課題を解決するための手段】上記目的を達成するために、本発明は、机上の指先の画像を用いるマンマシーン・インターフェース・システムであって、机上の画像を得る赤外線カメラと、前記赤外線カメラによる画像を入力して、処理する処理システムとを有しており、前記処理システムは、画像から手領域を体温を利用して抽出し、抽出した手領域から、指先を予め用意したパターンとのパターン・マッチングにより特定し、特定した指先の机上の座標を求めることを特徴とするマンマシーン・インターフェース・システムである。前記予め用意したパターンは、円形のパターンとすることもできる。これにより、指先の認識精度を高め、しかも高速に認識処理を行うことができる。

【0007】さらに、前記求めた指先が1本であるときは、指差し動作であると認識することができ、この指差し動作で、マウス等で行うポインティング操作を行うことができる。さらに、撮影位置の制御ができるカメラを備えとともに、前記指差し動作を行うことで、その指先の机上の座標に、前記カメラの撮影位置を合わせる制御を行うことができる。これにより、前記カメラで撮影した画像を取り込み、取り込んだ画像から、バーコードを抽出して、バーコード認識処理を行うこともできる。さらに、音声入力手段を備えとともに、前記音声入力手段による音声を入力し、音声認識処理を行うことにより、音声によるコマンド等を入力することができる。さ

らに、 프로젝タを備え、前記 프로젝タにより、コンピュータ画面を机上に投影することにより、コンピュータ画面とのインタラクティブな処理も行うことができる。上述の処理を行わせるプログラムを格納した記録媒体も本発明である。

【0008】

【発明の実施の形態】本発明のマンマシーン・インターフェース・システムは、以下の特徴を有しているものである。

- (1) 赤外線カメラを利用して、ユーザの皮膚領域からの放射光を測定することにより、ユーザの手領域を安定に抽出する。
- (2) 正規化相関等に基づいた、テンプレート・マッチングの利用により、ユーザの指先の高速追跡を実現する。
- (3) 机上の小さなバーコードの認識を行なうため、ズーム機能付パンチルト・カメラを用いてユーザの指先周辺を拡大追跡する。
- (4) 処理の高速化のため、画像処理ハードウェアの利用と分散処理の適用を行う。

以下に、これらの特徴を中心として、本発明の実施形態を、図面を参照して、詳細に説明する。

【0009】図1は、本発明の実施形態であるエンハンスド・デスク(Enhanced Desk)の外観図である。図1に示すように、通常の机10の上方に 프로젝タ30を取り付け、コンピュータ(図示せず)の画面を鏡44で反射させ、机上に画面12を投影する。机上を観察するカメラは2台設置した。1台は赤外線カメラ50であり、机10左側から上を向けて設置し、表面反射鏡42ごとに机上の撮影を行なう。もう1台のズーム機能付パンチルト・カメラ20は、机の右上方に設置し、机上の一部分を拡大して撮影することができる。また、図示していないが、スピーカとマイクも設置されており、音声等の入出力の機能も持っている。

【0010】図2は、実施形態におけるシステム構成図を示している。図2において、コンピュータ・システム60で生成された投影画像が、 프로젝タ30により表面反射鏡44を介して、机10上に投影される。また、机10上の画像を撮像しているパンチルト・カメラ20の姿勢等は、コンピュータ・システム60により制御されている。パンチルト・カメラ20で撮像された画像も、コンピュータ・システム60内に取り込まれて処理される。赤外線カメラ50で撮像された机上の画像も、コンピュータ・システム60に取り込まれて処理される。本発明のエンハンスド・デスクでの主な処理の流れを以下に示す。

- ① ユーザの手領域を、赤外線カメラ50の画像より抽出する。
- ② 抽出された手領域からユーザの指先点を検出する。
- ③ カメラ画像上の指先点の位置座標から、キャリブレ

(4)

特開2001-282456

5

6

ーションによりあらかじめ求めておいた射影変換パラメータを用いて、机上の位置座標に変換を行なう。

④ 机上の指先位置情報をもとに机上に投影したオブジェクトや情報とインタラクションをとる。

これらの処理について、以下詳細に説明する。

【0011】<手領域の抽出>手領域の抽出には、背景差分やカラー・カメラによる肌色抽出が一般的に用いられている。しかし、これらの手法では、背景の状態によっては手領域を抽出することが困難な場合が多い。特に本システムで想定される作業環境では、机上に紙書類や書籍などを置くことが考えられ、プロジェクトにより電子情報を投影するために、ユーザの手領域の色が一定ではないなどの問題点がある。そこで、本システムでは、手領域を安定に抽出するために赤外線カメラ50を利用した。机上を観察している赤外線カメラの画像から、人の体温をもとに手領域を切り出すことができる。赤外線カメラを利用して、人の体温の近傍（30℃～34℃）を撮影することにより、机上の手の様子を例えばNTSCの256階調の映像として得ることができる。この映像を一定の閾値で2値化処理し手領域を抽出する。図3に、赤外線カメラからの画像をもとに、手領域の抽出処理を説明する。図3（a）は、赤外線カメラ50からの画像である。コンピュータ・システム60で取り込む際の解像度は256×220画素である。そして、この画像から、人の体温の近傍（30℃～34℃）の画像を抽出する。これを示したのが図3（b）であり、手領域の部分抽出して2値化したものである。図3（b）に示すように、この手法により、背景や照明の変化に影響されず手領域だけを安定に抽出することができる。赤外線カメラは、例えばニコン・サーマルビジョン（LAI RD3A）を使用することができる。このカメラは、有効画素数768×465画素、フィールドタイム1/60秒で-20℃～50℃までの範囲を撮影することができる。

【0012】<指先位置の認識>指先位置の検出は、ユーザの手は机上にあり、カメラとユーザの手の距離はほぼ一定で見かけ上の大きさが極端に変わらないことから、一定の大きさのテンプレートで指先位置を検出することができる。指先位置の認識は一般に、指先形状の輪郭が円に近いことに基づき、円形テンプレートを用いて、手領域周辺で正規化相関に基づくパターン・マッチ

ングを行なうことで検出する。ユーザの手は机上にあり、見かけ上の大きさが極端に変わらないことから、パターン・マッチング処理の対象を手領域の一定範囲内のみに限定し、処理の高速化を図っている。この指先位置の認識を図4を用いて説明する。パターン・マッチング処理は、例えば手領域を切り出して2値化した256×220画素の画像（図3（b）参照）に対して、図4（a）に示す15×15画素の大きさのテンプレートを用いて、相関値を用いて一致検出を行う。パターン・マッチング処理の後、あまりにも近接した場所でも多くの相関値の高い点が見つかった場合、そのうちの一番相関値の高い点を残し、残りを指先候補点から除外する。これによりいくつかの指先候補点を得られるので、それらのうち指先点以外の誤認識された点を除外する。この処理は、図4（b）に示すように、テンプレートの周囲8点（矩形の4頂点及び各辺の二等分点）に対応するピクセルを調べ、指先点かどうかを判定する。この処理により指先がテンプレートを突き抜けていたり（図4（b）のERROR）、テンプレートの周囲に指らしきピクセルが何もない場合なども指先候補点から除外する。最終的に、図4（c）に示すように、指先候補点を相関値の高い順に5つまで採用しユーザの指先点とする。本システムにおける指先検出処理の実施例としては、日立画像処理ボードIP5010を用いて行なっている。IP5010は白黒濃淡画像で40画面分のメモリをもった画像処理ボードであり、メモリ上に保存した画像間で2値化や差分といった画像処理を、ハードウェアで高速に行なうことのできるボードである。OSはWindows NT（登録商標）およびLinuxに対応しており、多くの画像処理関数がライブラリとして用意されている。本実施例では、ネットワークとの親和性を鑑み、OSにLinuxを利用した。

【0013】<カメラ画像座標と机上平面座標の補正>赤外線画像中で検出された位置が机上でどの位置に対応するかを知るため、赤外線カメラ画像座標と机上平面座標の補正を行なう。ここでは、赤外線カメラ画像座標と机上平面座標の対応点をいくつか定義し、その関係を元に赤外線カメラ画像座標(x, y)と机上平面座標(x', y')の間の対応を式(1)のように射影変換を用いて表現する。

【数1】

$$\begin{cases} x' = c_1x + c_2y + c_3xy + c_4 \\ y' = c_5x + c_6y + c_7xy + c_8 \end{cases} \quad (1)$$

対応点を求めるために、赤外線カメラ対応点の測定を説明する図5を用いて説明する。机10に対して、机上平面の座標系を設定する。このために、机上の目標点の位置にキャリブレーション板70を作成し、これを机上に

置くことで机上座標系を設定する。手領域を切り出す赤外線カメラでは、温度差が画像として出力されるため、図5に示すように、このキャリブレーション板の対応点に小型電球72を埋め込み、点灯した電球を赤外線カメ

(5)

特開2001-282456

7

8

ラで撮影して対応点を計測する。対応点の組を計測後、その関係をもとに射影変換パラメータ( $c_1 \sim c_6$ )を決定する。このパラメータを用いて変換を行なうことによって画像のゆがみや、位置のずれなどを補正することができる。この射影変換パラメータを得るためには最低4組の対応点があればよいが、本システムでは、安定に解を求めるため、図5に示すように対応点(小型磁球)を9組用意した。この9組に対して、上記の連立方程式を特異値分解法を用いて解くことで、変換のパラメータを求めた。本システムでは、処理軽減のため机上の座標系とプロジェクタで机上に投影した画像の座標系を同一とし、机上面座標の値とプロジェクタ投影画像の画素の位置座標を等しくしている。これにより、机上面上の指先位置にプロジェクタの画像を投影するには、単に机上面座標上での座標値にオブジェクトを描画した画像を投影すればよく、処理の軽減を図ることができる。

【0014】<指先周辺の注視によるバーコード認識>本システムでは、画像処理を行なって認識したユーザの指先が1つだった場合、ユーザが指差し動作をしていると判定することもできる。このことにより、例えば入差し指だけで指すことにより、指差し動作であると認識して、この指先の座標を得ることにより、マウス等のポインティングの代わりに用いることができる。この指差し動作をしている指先の周辺が注目している領域であるとみなすこともできる。この応用であるパンチルト・カメラ20で、1本の指先の注目領域を追跡する処理を、以下に説明する。

【0015】(指先追跡処理の流れ)パンチルト・カメラ20における指先追跡の処理の流れを以下に示す。

- ① 机上面とパンチルト駆動平面の2平面間で対応点を計測し、射影変換パラメータを算出しておく。
- ② 画像処理によって赤外線カメラ画像上の指先点の位置を計測する。
- ③ 赤外線カメラ画像上の指先点の位置を射影変換により机上面上の座標に変換する。
- ④ 机上の指先点の位置をパンチルト駆動平面に射影変換を用いて変換する。
- ⑤ パンチルト駆動平面上の座標をカメラにVISCAコードで送出し、カメラをユーザの指先点に向かせる。
- ⑥ 指先点に向けたカメラの画像を処理する。

【0016】(射影変換パラメータの算出)本システムでは、机上の指先位置が、カメラ20のパンチルト駆動平面上でどの座標位置にあるかがわかればよい。この処理は、パンチルト・カメラ20が現在、駆動平面上でどの座標値をとっているかを知ることができるので、これを利用して行う。先に、カメラ画像座標と机上座標の補正を行なったことを説明したが、指先追跡においても同様の手法を用いて、カメラ20による座標補正を行っている。そのためまず、今までの処理で指先の机上座標はわかっているため、その机上位置に対応したパンチ

ルト・カメラ20の駆動平面上的位置を算出する。対応点は4組以上計測する。この計測は、机上の対応点をパンチルトカメラが中心に捕らえるように位置を調節し、その時点でのパンチルト駆動平面上的座標値を対応点として測定する。この対応点をもとに先の射影変換を用いて、机上面座標からパンチルト駆動平面上的座標に変換するパラメータを求める。指先追跡のためのカメラ20は、座標を指定することでその方向にカメラを向けることができる機能を有している。この機能を利用して、ユーザの指先を追跡させている。実施例として、指先追跡のための追跡カメラ20は例えば、垂直方向-7794~7794、水平方向-15570~15570の大きさのパンチルト駆動平面をもっているSony EVI-Q20パンチルトカメラを使用し、PCのシリアルポートからVISCAコマンドをカメラに送信することによって、カメラのパンチルト制御を行なっている。

【0017】(指先追跡処理)さて、指先追跡処理を説明する。赤外線カメラ50からの画像を上記の画像処理によって、机10上の指先点の位置を計測する。赤外線カメラ画像上の指先点の位置を射影変換により机上面上の座標に変換するとともに、1本の指先のみであることを認識する。求めた机上の指先点の位置を、上述のパラメータにより、カメラ20のパンチルト駆動平面に射影変換を用いて変換する。パンチルト駆動平面上的座標をカメラにVISCAコードで送出し、カメラ20をユーザの指先点に向かせる。これにより、指先点に向けたカメラ20の画像を取り込むことができる。

【0018】(バーコードの認識)以下に、このカメラ20の映像を用いて、指先の拡大追跡を行ない、バーコード等の小さい物体の認識を行なっている例を、図6を用いて説明する。図6は、指先の拡大追跡を行ない、バーコード認識を行っている様子を示す。現実物体(紙書類、書籍等)に、バーコードを取り付けることにより認識を行ない、電子情報とのリンクを作成することができる。バーコードには二次元マトリックス・コードを利用した。この二次元バーコードはコードの種類だけでなく、コードのある位置や、向きなどもコードから得ることができるので、物体のコードが貼付されている位置を記憶していれば、コードから物体の姿勢を計算することができる。図6(a)に、二次元バーコードの例を示す。この様なバーコードは、パンチルト・カメラ20で撮影した画像の中に存在すれば認識される。認識するためにはカメラ画像中である程度の大きさが必要であるが、パンチルト・カメラ20が拡大した映像を撮影しているので、机上面においた書類に貼り付けたバーコードでも、約1.5×1.5cmの大きさのものまで認識することが可能である。机上面ではユーザの指差し動作によって、バーコードが指差されれば、そのコードを認識し、それに対応したインタラクションを起こすことができる。図6

(b)に、このバーコードに対する、指先の拡大追跡を

(6)

特開2001-282456

9

10

行なった様子を示す。図6(b)は、カメラ20の画像を2値化した画像を示している。まず、赤外線カメラ50からの画像上の指先点の位置を画像処理によって計測する。これにより、1本の指で指していること、およびその指先点の位置座標をえることができる。次に、赤外線カメラ画像上の指先点の位置を射影変換により机上平面上の座標に変換する。得られた机上の指先点の位置を、パンチルト駆動平面に射影変換を用いて変換して、得られたパンチルト駆動平面上の座標をカメラ20にVI

10

20

SCAコードで送出し、カメラ20をユーザの指先点に向かせる。そして、指先点に向けたカメラ20の画像(図6(b))を処理して、バーコード画像を認識して、バーコードの読み取り処理を行う。バーコードの認識処理は実施例として、例えば、ビデオ映像をSGI社のシステムであるO2のビデオ入力端子から320×240画素の大きさで取り込み、2値化等の処理を行ない、バーコード認識ライブラリを用いて認識を行なうことができる。画像入力はSGI社のビデオ・ライブラリを使用している。また、バーコード認識処理はソフトウェアで行なっており、10〜15f/s程度の速度で実行が可能である。

【0019】<音声認識>本システムには、ユーザとのインタラクションの補助的役割として音声認識を使用している。ここではユーザの指差し動作やオブジェクトの移動や回転といった操作モードの変更を行なうためのキーワードを認識させるためにのみ利用している。音声認識はIBM ViaVoiceの音声認識エンジンを使用して行なっている。本システムでは、あらかじめ登録してある単語をユーザが発声したときに何らかのインタラクションを起こすために、音声認識プロセスはメッセージ・サーバに対して、認識した単語の種類をメッセージとして宛に送っている。サーバに送られたメッセージは、情報提示プロセスが受け取り、その単語に対応したインタラクションを机上で起こすことが可能である。

30

40

50

オ入力に取り込み、ソフトウェア処理によって2次元バーコードの認識を行なっている。認識結果は宛にメッセージ・サーバ68に送出される。音声認識は、例えば、PentiumII 330Mhzのパソコン62にWindowsNTとViaVoiceを導入して処理を行なっている。ユーザがマイクに向かって話した音声すべてを認識させ、そのうちあらかじめ登録してある単語を認識した場合に、特定のメッセージをメッセージ・サーバ68に送出している。最後に、宛にメッセージ・サーバ68に貯えられている情報を取り出し、それに対応するインタラクションや画像を作成する情報表示プロセスがある。これはいわばアプリケーション・プロセスであり、SGI O2システム66で処理を行なっている。各々の処理間の通信を行うメッセージ・サーバ68には、ダブル空間通信システムTS System/s、通称Lindaを用いている。このシステム68はネットワーク上でダブル・スペースと呼ばれる空間を共有し、その空間を介して任意の文字列をセットとしたダブルと呼ばれるメッセージをやり取りすることによって、通信を実現している。非同期通信であるため、個々のプロセスが完全に独立して動作することができ、また新しいプロセスの追加や変更が容易であるという利点がある。本システムでは、指先を毎秒20フレーム以上の実用的な速度で、安定して検出することができた。

【0021】<アプリケーション例>

(簡易文字認識) 机上でのユーザの指差し動作を追跡することを利用して、ユーザが指で書いた文字を認識することができる。例えば、プロジェクト30から机上に投影されたマス目に対して指で数字を書き入れると、その数字を認識して、書いた数字を机上にプロジェクト30により表示することができる。ユーザの指先に単純なポインティング以外の意味を持たせるために利用できる。(その他の応用) その他の応用として、ユーザの指先の位置情報だけでなく、5本の指先の軌跡を利用したジェスチャ認識等も可能である。カメラによる拡大撮影が可能になったことから、紙面の小さい文字も認識できると思われるため、文字認識を利用し、紙書類と電子情報の統合的な利用のための機能を付加することができる。また、本システムを遠隔協調作業への応用として、遠隔地の2つの机上を共有することによる机上平面を媒介としたコミュニケーション環境の構築もできる。本発明は、上述の分散コンピュータ・システムばかりではなく、複数のシステムから構成されるクライアント・サーバ・システムやスタンド・アローンのシステムに適用してもよい。本発明に関するプログラムを格納した記憶媒体から、プログラムをシステムで読み出して実行することにより、本発明の構成を実現することができる。この記録媒体には、フロッピー(登録商標)・ディスク、CD-ROM、磁気テープ、ROMカセット等がある。

【0022】

【発明の効果】 上述のように、本発明を用いると、机上

(7)

特開2001-282456

11

12

での電子情報と現実物体を統合的に扱うことのできる机「Enhanced Desk」実現のために必要であるインターフェースが実現できる。また、カメラで指先を拡大追跡しているため、例えば、小型のバーコードの認識が可能となり、机上の物体に不自然に大きなバーコードをつける必要がなくなった。ユーザは、本発明のインターフェースを用いることにより、例えば机上に投影された3Dオブジェクトを指先で回転させながら移動したり、書籍に添付したバーコードを指差すことで関連するホームページを表示させるなど、従来のマウスやキーボードよりも、直感的で現実と密接にリンクしたインタラクションを電子情報との間で行なうことができる。

【図面の簡単な説明】

【図1】エンハンスド・デスクの外観図である。

【図2】本発明のシステム構成図である。

【図3】手領域の抽出を示す図である。

【図4】指先認識を説明する図である。

【図5】対応点の測定を説明する図である。

\*【図6】指先の拡大追跡を説明する図である。

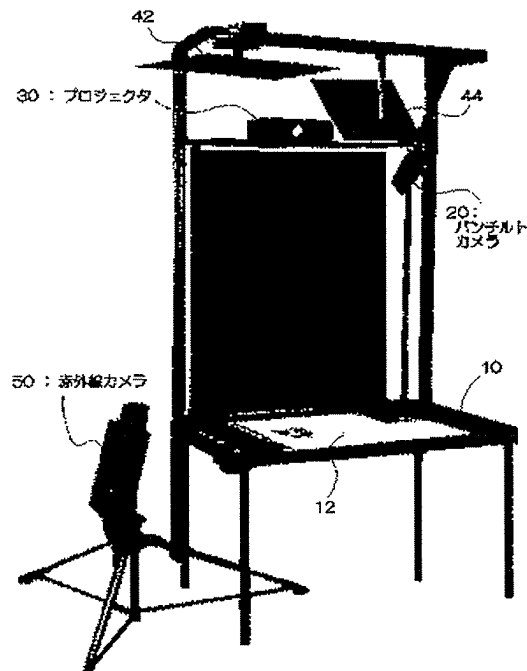
【図7】分散処理を説明する図である。

【符号の説明】

10	机
12	画面
20	ズーム機能付パンチルト・カメラ
30	プロジェクタ
42	表面反射鏡
44	表面反射鏡
50	赤外線カメラ
60	コンピュータ・システム
62	パソコン
64	パソコン
66	SGI O2・システム
68	メッセージ・サーバ
70	キャリブレーション板
72	小型電球

\*

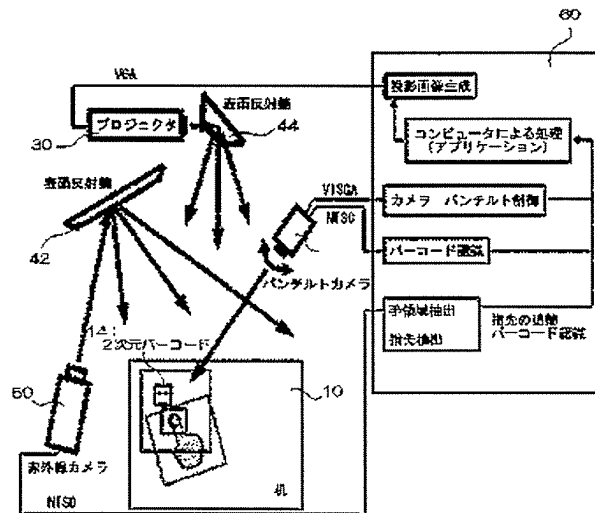
【図1】



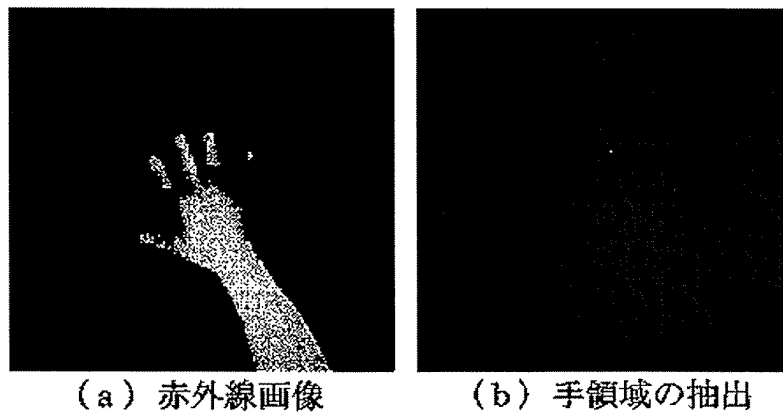
(8)

特開2001-282456

【図2】



【図3】

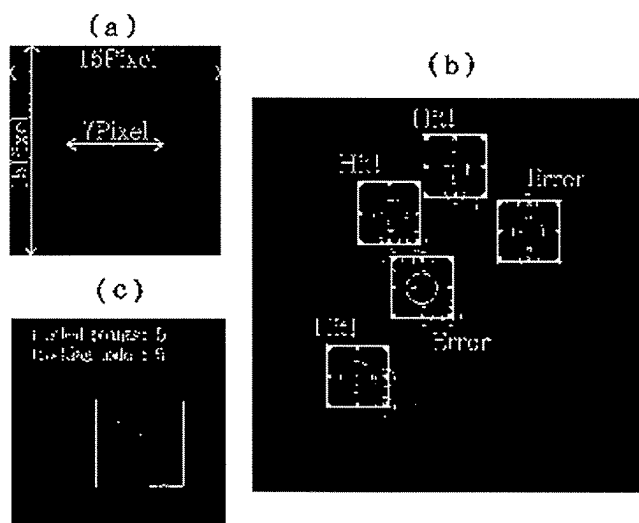




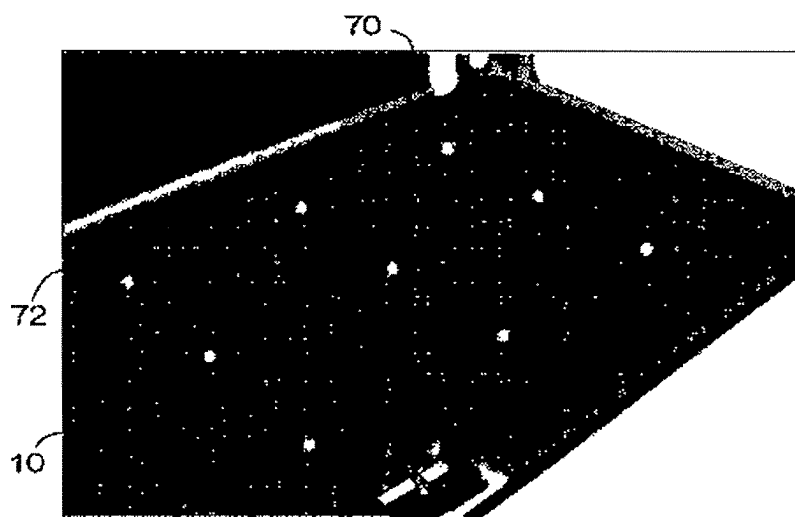
(9)

特開2001-282456

【図4】



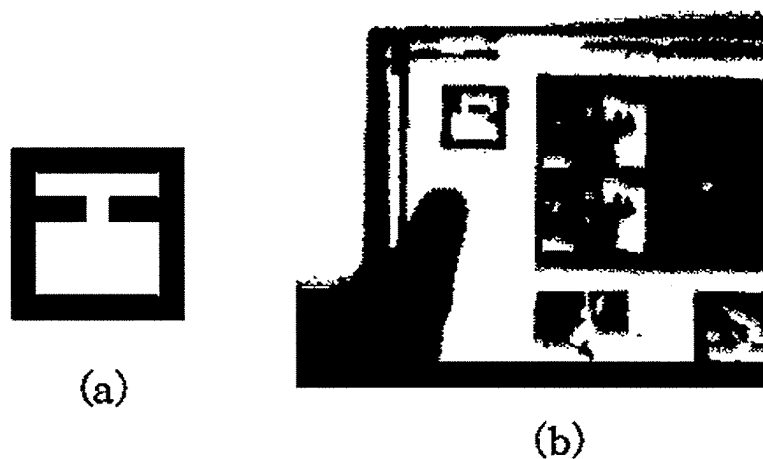
【図5】



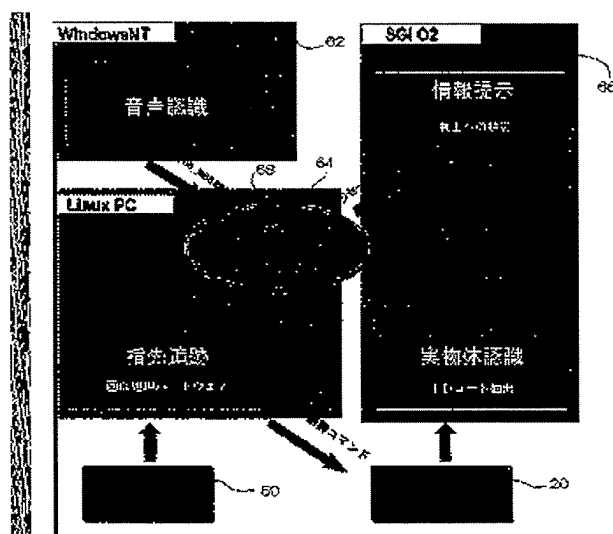
(10)

特開2001-282456

【図6】



【図7】



フロントページの続き

(51)Int.Cl.	識別記号	F i	リマート(参考)
G 0 6 T 1/00	3 4 0	G 0 6 T 1/00	3 4 0 Z 5 L 0 9 6
7/00	3 0 0	7/00	3 0 0 D

(11)

特開2001-282456

Fターム(参考) 5B057 AA20 BA08 BA19 CA08 CA12  
CA16 CB06 CB12 CB16 CC01  
CE09 CE12 DA08 DB02 DB08  
DC09 DC34  
5B068 AA05 AA36 BB18 BC05 BD09  
CC11 CC19 DD11  
5B072 CC24 DD02  
5B087 AA10 AB02 AD02 AE06 BC06  
BC32 CC09 CC33  
5E501 AA01 BA05 CA01 CB14 CB15  
CC03 CC04 CC11 CC14 DA13  
DA16 EA03 EA13 EA21 EB06  
5L096 AA07 BA08 BA18 CA18 DA02  
EA35 EA43 GA17 HA08 JA09  
JA28 LA01